The Annals of Mathematical Statistics
1968, Vol. 39, No. 4, 1310–1315

THE ϵ-ENTROPY OF CERTAIN MEASURES ON [0, 1][1]

By T. S. Pitcher

University of Southern California

1. Introduction. The epsilon-entropy of a probability distribution on a metric space was introduced in [4]. If C is a countable covering of the space by measurable sets we write $\|C\| = \max_{A \varepsilon C}$ (diameter (A)), $\#(C)$ = number of sets in C and

$$H(C) = \sum_{A \varepsilon C} P(A) \log (P(A))^{-1}.$$

Then the epsilon-entropy H_ϵ is given by

$$H_\epsilon = \inf_{\|C\| \leqq \epsilon} H(C).$$

In this paper we derive estimates of the asymptotic behavior of H_ϵ for certain singular measures on [0, 1]. The metric will be the usual length and we will write $|A|$ for the length of an interval A.

It will be convenient to use the notation $\phi(x) = x \log 1/x$. The function ϕ is convex and has the property that if $p_i \geqq 0$, $\sum_1^n p_i = 1$ then $\sum_1^n \phi(p_i) \leqq \log n$.

The theorems of this paper give asymptotic comparisons of H_ϵ with $\log \epsilon^{-1}$ which is approximately the ϵ-entropy of Lebesgue measure on [0, 1]. The asymptotic ratios are given in terms of various information theoretic quantities.

2. Measures related to N-adic expansions. Let N be a fixed integer, $N \geqq 2$ and let $(a_i, i = 1, 2, \cdots)$ be a stationary ergodic stochastic process taking the values $0, 1, \cdots, N - 1$. We assume that no fixed sequence (a_i^0), $i = 1, 2, \cdots$, has positive probability. Define $k_i(x)$ for irrational x in [0, 1] by

$$x = \sum_{i=1}^{\infty} k_i(x) N^i$$

where the sum on the right is the N-adic expansion of x. Write

$$I_n(l_1, \cdots, l_n) = [x \mid k_1(x) = l_1, \cdots, k_n(x) = l_n],$$
$$I_n(x) = I_n(k_1(x), \cdots, k_n(x)).$$

The probability measure P associated with the process induces a measure, which we also call P, on [0, 1] through the formula

$$P(I_n(l_1, \cdots, l_n)) = P(a_1 = l_1, \cdots, a_n = l_n).$$

According to the Shannon-Macmillan-Breiman theorem

$$\lim_{n \to \infty} n^{-1} \log P(I_n(x)) = -h(P) \text{ a.e. } (P)$$

where $h(P)$ is the entropy of the shift operator.

Received 20 November 1967.

[1] This research was partially supported by National Science Foundation Grant GP-6216. Part of this work was done while the author was a consultant at the Jet Propulsion Laboratory, Pasadena, California.

Since we are going to work with coverings by intervals we will need the following lemma.

LEMMA 2.1. *If we set*

$$H_\epsilon^I = \inf [H(C) \mid \|C\| \leqq \epsilon \quad \text{and} \quad C \quad \text{is a covering by intervals}]$$

and if $\epsilon = n^{-1}$ *for some* n *then*

$$\tfrac{1}{2}H_\epsilon^I \leqq H_\epsilon \leqq H_\epsilon^I.$$

PROOF. The second inequality is obvious. Let D_n be the interval covering $[[k/n, (k+1)/n] \mid k = 0, \cdots, n-1]$ and let C be an arbitrary covering with $\|C\| \leqq n^{-1}$. Let $A_1^k, \cdots, A_m^k$ be the sets of C which intersect $I_k = [k/n, (k+1)/n]$. Then

$$\begin{aligned}\phi(P(I_k)) &\leqq \phi\left(\textstyle\sum_{j=1}^m P(A_j^k)\right) \\ &\leqq \textstyle\sum_{j=1}^m \phi(P(A_j^k)).\end{aligned}$$

Summing on k and noting that each A can hit at most two I_k's we have $H(D_n) \leqq 2H(C)$ and the result follows.

Let C_n be the covering of [0, 1] by N-adic intervals, i.e., $C_n = \{J_k = [kN^{-n}, (k+1)N^{-n}), k = 0, 1, \cdots, N^n - 1\}$.

LEMMA 2.2. *For* $\epsilon = N^{-n}$,

$$\tfrac{1}{2}H(C_n) \leqq H_\epsilon^I \leqq H(C_n).$$

PROOF. The second inequality is obvious from the definition of H_ϵ^I, since $\|C_n\| = N^{-n}$. Let $C = \{I_i\}$ be any covering by intervals with $\|C\| \leqq \epsilon$. If $I_i^k, \cdots, I_{i_m}^k$ are the intervals intersecting J_k then

$$\textstyle\sum_{j=1}^m \phi(P(I_{i_1}^k)) \geqq \phi\left(\sum_{j=1}^m P(I_{i_1}^k)\right) \geqq \phi(P(J_k)).$$

Summing on k and noting that each I_i hits at most two J_k's we have $2H(C) \geqq H(C_n)$ from which the result follows.

Now fix $\delta > 0$ and set

$$C_n' = [J_k \mid |n^{-1} \log P(J_k) + h(P)| \leqq \delta],$$

$$C_n'' = [J_k \mid |n^{-1} \log P(J_k) + h(P)| > \delta],$$

$$H_n' = H(C_n'), \qquad H_n'' = H(C_n''), \qquad q(n, \delta) = P\left(\textstyle\bigcup_{J_k \varepsilon C_n''} J_k\right).$$

By the Shannon-Macmillan-Breiman theorem $q(n, \delta)$ goes to 0 as n goes to ∞.

LEMMA 2.3. $H_n'' \leqq q(n, \delta) n \log N + \phi(q(n, \delta))$.

PROOF. Set $q = q(n, \delta)$ and

$$\begin{aligned} P_k &= P(J_k)/q \quad &&\text{if} \quad J_k \varepsilon C_n'' \\ &= 0 &&\text{if} \quad J_k \varepsilon C_n'. \end{aligned}$$

Since $\sum P_k = 1$ we have

$$\begin{aligned} n \log N &\geqq \textstyle\sum_0^{N^n-1} \phi(P_k) \\ &= q^{-1} \textstyle\sum_{C_n''} \phi(P(J_k)) + q^{-1} \sum_{C_n''} P(J_k) \log q = q^{-1} H_n'' - \log q^{-1}. \end{aligned}$$

THEOREM 2.1.
$\frac{1}{4}h(P)/\log N \leqq \liminf_{\epsilon\to 0} H_\epsilon/\log \epsilon^{-1} \leqq \limsup_{\epsilon\to 0} H_\epsilon/\log \epsilon^{-1} \leqq h(P)/\log N.$

PROOF. With the notation as above,

$$(1 - q(n, \delta))n(h(P) - \delta) \leqq \sum_{C_n'} \phi(P(J_k)) = H_n' \leqq (1 - q(n, \delta))n(h(P) + \delta).$$

Thus, since $q(n, \delta) \to 0$,

$$\liminf_{n\to\infty} H(C_n)/n \log N \geqq \liminf_{n\to\infty} H_n'/n \log N \geqq (h(P) - \delta)/\log N$$

and

$$\begin{aligned}\limsup_{n\to\infty} H(C_n)/n \log N &= \limsup_{n\to\infty} (H_n' + H_n'')/n \log N \\ &\leqq \limsup_{n\to\infty} [q(n, \delta)n \log N + \phi(q(n, \delta)) \\ &\quad + (1 - q(n, \delta))n(H + \delta)](n \log N)^{-1} \\ &= [h(P) + \delta]/\log N.\end{aligned}$$

Since δ is arbitrarily small this proves

$$\tfrac{1}{2}h(P)/\log N \leqq \liminf_{\epsilon\to 0} H_\epsilon^I/\log \epsilon^{-1} \leqq \limsup_{\epsilon\to 0} H_\epsilon^I/\log \epsilon^{-1} \leqq h(P)/\log N$$

and an application of Lemma 2.1 proves the theorem for the sequence $\epsilon_n = N^{-n}$. In general if we define $n(\epsilon)$ by

$$N^{-n(\epsilon)-1} \leqq \epsilon < N^{-n(\epsilon)}$$

then

$$\begin{aligned}\liminf_{\epsilon\to 0} H_\epsilon/\log \epsilon^{-1} &\geqq \liminf_{\epsilon\to 0} [H_{N^{-n(\epsilon)}}]/(n(\epsilon) + 1) \log N \\ &= \liminf_{n\to\infty} H_{N^{-n}}/(n + 1) \log N \geqq \tfrac{1}{4}h(P)/\log N\end{aligned}$$

and

$$\begin{aligned}\limsup_{\epsilon\to 0} H_\epsilon/\log \epsilon^{-1} &\leqq \limsup_{\epsilon\to 0} H_{N^{-n(\epsilon)-1}}/n(\epsilon) \log N \\ &= \limsup_{n\to\infty} H_{N^{-n-1}}/n \log N \leqq h(P)/\log N.\end{aligned}$$

3. Measures related to continued fraction expansions. Every irrational number in [0, 1] has a unique infinite continued fraction expansion

$$x \doteq \cfrac{1}{a_1(x) + \cfrac{1}{a_2(x) + \cfrac{1}{\ddots}}}$$

where the $a_i(x)$ are positive integers. We will write

$$I_n(k_1, \cdots, k_n) = [x \mid a_1(x) = k_1, \cdots, a_n(x) = k_n]$$

and

$$I_n(x) = I_n(a_1(x), \cdots, a_n(x)).$$

If $(a_i, i = 1, 2, \cdots)$ is a stationary ergodic process taking positive integer values then the probability measure P induces, as before, a measure P on [0, 1] such that

$$P(I_n(k_1, \cdots, k_n)) = P(a_1 = k_1, \cdots, a_n = k_n).$$

As before we assume that the induced measure has no atoms. We assume that $\sum_{i=1}^{\infty} \phi(P(a_1 = i)) < \infty$ so that $h(P) < \infty$ and that $h_0(P) = 2\int_0^1 \log t^{-1} P(dt) < \infty$. Then, by Chung's extension of the Shannon-Macmillan-Breiman theorem [1],

$$\lim_{n\to\infty} n^{-1} \log P(I_n(x)) = -h(P) \text{ a.e. } (P)$$

and by Theorem 2.2 of [3] (with $f(x) = x^{-1}$)

$$\lim_{n\to\infty} n^{-1} \log |I_n(x)| = -h_0(P) \text{ a.e. } (P).$$

Fix $\delta > 0$ and set

$$C_n' = [I_n(k_1, \cdots, k_n) \mid |n^{-1} \log P(I_n(k_1, \cdots k_n)) + h(P)| < \delta \quad \text{and} \quad |n^{-1} \log |I_n(k_1, \cdots, k_n)| + h_0(P)| < \delta].$$

Then $\#(C_n') \leqq e^{n(h_0(P)+\delta)} + 1$ and $\|C_n'\| \leqq e^{-n(h_0(P)-\delta)}$. We can find a covering C_n'' of the remainder of [0, 1] with $\#(C_n'') \leqq 2(e^{n(h_0(P)+\delta)} + 1)$ and $\|C_n''\| \leqq e^{-n(h_0(P)-\delta)}$. In fact if we start with intervals of the form $[ke^{-n(h_0(P)-\delta)}, (k+1)e^{-n(h_0(P)-\delta)}]$ and successively delete the intervals of C_n' each deletion will add at most one interval and this will give the desired covering. Let C_n be the combined covering. For convenience we take n so large that $2(e^{n(h_0(P)+\delta)} + 1) \leqq e^{n(h_0(P)+2\delta)}$.

If we set $H_n' = H(C_n')$ and $H_n'' = H(C_n'')$ then we can prove exactly as in Lemma 2.3, that

$$H_n'' \leqq q(n, \delta)n(h_0(P) + 2\delta) + \phi(q(n, \delta))$$

where $q(n, \delta) = \sum_{I \epsilon C_n'} P(I)$ goes to 0 as n goes to ∞.

THEOREM 3.1.

$$\limsup_{\epsilon\to 0} H_\epsilon / \log \epsilon^{-1} \leqq h(P)/h_0(P).$$

PROOF. The proof is almost an exact duplicate of the corresponding part of the proof of Theorem 2.1.

It is not possible to get the opposite inequality by the same device as before since an interval of length $e^{-n(h_0(P)-\delta)}$ could hit roughly $e^{2n\delta}$ intervals of C_n'.

4. The case $h_0(P) = \infty$. In this section we are concerned with the case where $(a_i, i = 1, 2, \cdots)$ is an integer valued stationary ergodic process as in the previous section, with P the measure induced by the process through the continued fraction representation and where

$$\sum_{i=1}^{\infty} \phi(P(a_1 = i)) < \infty$$

so that $h(P) < \infty$, but

$$h_0(P) = 2\int_0^1 \log t^{-1} P(dt) = \infty.$$

To see that such cases exist, note first that $h_0(P) = \infty$ if and only if $\int \log a_1(t)P(dt) = \infty$ since $\log a_1(t) \leqq \log t^{-1} \leqq \log(1 + a_1(t)) \leqq 1 + \log a_1(t)$. Thus if we take the a_i to be independent with $P(a_1 = i) = p_i$ we need only choose the p_i so that $\sum_{i=1}^{\infty} \phi(p_i) < \infty$ while

$$\sum_{i=1}^{\infty} p_i \log i = \int_0^1 \log a_1(t) P(dt) = \infty.$$

We will need some facts about continued fraction expansions (see [2]). If we write

$$P_n(x)/Q_n(x) = \cfrac{1}{a_1(x) + \cfrac{1}{a_2(x) + \cfrac{1}{\ddots + \cfrac{1}{a_n(x)}}}}$$

then $P_n(x)$ and $Q_n(x)$ are generated by:

$$\begin{aligned} P_0(x) &= 0, \qquad Q_0(x) = 1, \\ P_{n+1}(x) &= a_{n+1}(x)P_n(x) + P_{n-1}(x), \\ Q_{n+1}(x) &= a_{n+1}(x)Q_n(x) + Q_{n-1}(x), \end{aligned}$$

and we have

$$|I_n(x)| = [Q_{n-1}(x)\{Q_n(x) + Q_{n-1}(x)\}]^{-1}.$$

LEMMA 4.1. *There exist numbers $A_N \uparrow \infty$ such that*

$$\liminf_{n\to\infty} n^{-1} \log |I_n(x)|^{-1} \geqq A_N \text{ a.e. } (P).$$

PROOF.

$$\begin{aligned} \liminf_{n\to\infty} n^{-1} \log |I_n(x)|^{-1} &\geqq \liminf_{n\to\infty} 2n^{-2} \log Q_{n-1}(x) \\ &\geqq \liminf_{n\to\infty} 2n^{-2} \sum_{i=1}^{n-1} \log a_i(x) \\ &\geqq \liminf_{n\to\infty} 2n^{-2} \sum_{i=1}^{n-1} \log (\min (a_i(x), N)) \\ &= 2\int_0^1 \log (\min (a_i(x), N))P(dx) = A_N. \end{aligned}$$

Now let, for fixed N and $\delta > 0$,

$$C_n' = [I_n(k_1, \cdots k_n) \mid |n^{-1} \log P(I_n(k_1, \cdots, k_n)) + h(P)| < \delta$$
$$\text{and} \quad n^{-1} \log |I_n(k_1, \cdots, k_n)|^{-1} \geqq A_N],$$
$$q(n, \delta, N) = 1 - P(\textstyle\bigcup_{C_n'} I_n).$$

By Chung's theorem and the lemma above $q(n, \delta, N)$ goes to 0 as n goes to ∞. $\|C_n'\| \leqq e^{-nA_N}$ and $\#(C_n') \leqq e^{n(h(P)+\delta)}$ since

$$1 \geqq \textstyle\sum_{C_n'} P(I_n) \geqq \#(C_n')e^{-n(h(P)+\delta)}.$$

As before we can find a covering C_n'' of the complement with $\|C_n''\| \leqq e^{-nA_N}$ and $\#(C_n'') \leqq e^{nA_N} + 1 + e^{n(h(P)+\delta)}$. We take C_n to be the combined covering and assume for convenience that N is so large that $e^{nA_N} + 1 + e^{n(h(P)+\delta)} \leqq e^{2nA_N}$. As before

$$H_n'' = H(C_n'') \leqq q(n, \delta, N) \log \#(C_n'') + \phi(q(n, \delta, N))$$
$$\leqq q(n, \delta, N)2nA_N + \phi(q(n, \delta, N))$$

and

$$H_n' = H(C_n') \leqq n(h(P) + \delta).$$

This gives, for $\epsilon = e^{-nA_N}$,

$$H_\epsilon/\log \epsilon^{-1} \leqq (H_n' + H_n'')/nA_N \leqq (h(P) + \delta)/A_N$$
$$+ 2q(n, \delta, N) + \phi(q(n, \delta, N))/nA_N.$$

We can now proceed in the usual way to get the following extension of Theorem 3.1.

THEOREM 4.1. *If the stationary ergodic process* $(a_i, i = 1, 2, \cdots)$ *has* $h(P) < \infty$ *but* $h_0(P) = \infty$ *then*

$$H_\epsilon = o(\log \epsilon^{-1}).$$

REFERENCES

[1] CHUNG, K. L. (1961). A note on the ergodic theorem of information theory. *Ann. Math. Statist.* **32** 612–614.

[2] KHINTCHINE, A. YA (1963). *Continued Fractions.* Noordhoff, The Netherlands.

[3] KINNEY, J. and PITCHER, T. (1966). The dimension of some sets defined in terms of f-expansions. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **4** 293–315.

[4] POSNER, E., RODEMICH, E. and RUMSEY, H. (1967). Epsilon entropy of stochastic processes. *Ann. Math. Statist.* **38** 1000–1020.